W1518

PATENT ABSTRACTS OF JAPAN

(11)Publication number : 09-251353
(43)Date of publication of application : 22.09.1997

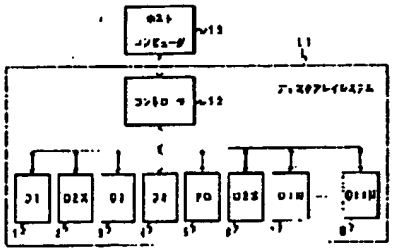
(51)Int.Cl. G06F 3/06
G06F 3/06
G11B 20/18
G11B 20/18

(21)Application number : 08-057719
(22)Date of filing : 14.03.1996

(71)Applicant : TOSHIBA CORP
(72)Inventor : TAKEKADO SHIGERU
INOUE TETSUO

(54) DISK ARRAY SYSTEM

(57)Abstract:
PROBLEM TO BE SOLVED: To provide a disk array system high in reliability and capable of being used for a long time exceeding the service life period of constituting respective disk drives independently of the service life of respective HDDs(hard disk drives) with the minimum increase of cost.
SOLUTION: This disk array system 11 incorporating the plural HDDs 1-10 is provided with a controller 12 for turning at least one or more disk drives to a stop state as standby disk drives, monitoring time lapse from the start of the stop state and shifting the disk drives of the stop state to an idling operation state every time prescribed time elapses. The controller 12 executes control for exchanging a failed disk drive HDD 2 and the standby disk drive 6. By such a system, the standby disk drives 6-10 exchangeable with the disk drive 2 with no residual life are prepared, and the standby disk drives 6-10 are made into the idling operation state at prescribed intervals, so that it is prevented from that a function becomes incomplete at the time of use.



LEGAL STATUS

[Date of request for examination]

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number]

[Date of registration]

[Number of appeal against examiner's decision

of rejection]
[Date of requesting appeal against examiner's decision of rejection]
[Date of extinction of right]

(19)日本国特許庁（JP）　(12)公開特許公報(A)　(11)特許出願公開番号

特開平9－251353

(43)公開日　平成9年(1997)9月22日

(51)Int.Cl.⁶	識別記号	庁内整理番号	FI		技術表示箇所
G06F 3/06	540		G06F 3/06	540	
	306			306H	
G11B 20/18	570		G11B 20/18	570Z	
	572			572F	

審査請求　未請求　請求項の数8　OL　（全 9 頁）

(21)出願番号　特願平8－57719

(22)出願日　平成8年(1996)3月14日

(71)出願人　000003078
株式会社東芝
神奈川県川崎市幸区堀川町72番地

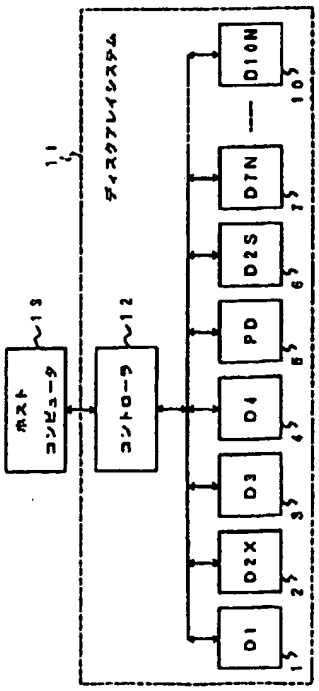
(72)発明者　竹門　茂
神奈川県川崎市幸区小向東芝町1番地　株式会社東芝研究開発センター内

(72)発明者　井上　徹夫
東京都港区芝浦一丁目1番1号　株式会社東芝本社事務所内

(74)代理人　弁理士　鈴江　武彦

(54)【発明の名称】　ディスク・アレイ・システム

(57)【要約】
【課題】構成する各ディスクドライブの寿命期間を越える長期間の使用を可能にして、最小限のコスト負担で各HDDの寿命に依存しない高信頼性のディスク・アレイ・システムを提供することにある。
【解決手段】複数のHDD1～10を内蔵するディスク・アレイ・システムであって、少なくとも1台以上のディスクドライブを予備ディスクドライブとして停止状態にして、停止状態の開始から時間経過を監視し、所定の時間が経過する毎に停止状態のディスクドライブをアイドリング動作状態に移行させるコントローラ12を備えたシステムである。コントローラ12は、故障したHDD2と予備ディスクドライブ6とを交換する制御を実行する。このようなシステムにより、寿命がきたディスクドライブ2と交換可能な予備ディスクドライブ6～10を準備すると共に、予備ディスクドライブ6～10を所定の間隔でアイドリング動作状態にすることにより、使用時に機能不全となるような事態を防止することができる。



【特許請求の範囲】

【請求項1】 複数のディスクドライブを内蔵して、各ディスクドライブを並列に駆動制御するディスク・アレイ・システムであって、

前記各ディスクドライブの中で、少なくとも1台以上のディスクドライブを予備ディスクドライブとして停止状態に移行させて、

前記停止状態の開始から時間経過を監視し、所定の時間が経過する毎に前記停止状態のディスクドライブをアイドリング動作状態に移行し、

所定の条件で動作状態のディスクドライブと前記予備ディスクドライブとを交換する制御を実行するコントローラ手段を具備したことを特徴とするディスク・アレイ・システム。

【請求項2】 前記コントローラ手段は、動作中のディスクドライブの中で故障したディスクドライブを停止状態に移行させて、

前記予備ディスクドライブを動作状態にすると共に、前記故障したディスクドライブに格納されている全データを復元して、前記予備ディスクドライブに格納する処理を実行することを特徴とする請求項1記載のディスク・アレイ・システム。

【請求項3】 前記コントローラ手段は、前記アイドリング動作状態として、前記予備ディスクドライブのディスクを定常回転速度で回転させる動作またはヘッドアクチュエータをディスクの半径方向に所定回数だけシークさせる動作の中で、少なくともいずれかの動作を実行させることを特徴とする請求項1記載のディスク・アレイ・システム。

【請求項4】 複数のディスクドライブを内蔵して、各ディスクドライブを並列に駆動制御するディスク・アレイ・システムであって、

前記各ディスクドライブの中で、少なくとも1台以上のディスクドライブを予備ドライブとして停止状態にし、

前記各ディスクドライブの連続動作時間を監視し、所定の時間が経過したディスクドライブを停止状態に移行させる処理を実行し、

前記停止状態に移行させるときに、前記予備ディスクドライブを動作状態に移行すると共に、前記停止状態に移行させるディスクドライブに格納されている全データを前記予備ディスクドライブに復元する処理を実行するコントローラ手段を具備したことを特徴とするディスク・アレイ・システム。

【請求項5】 N（≧3）台以上のディスクドライブを内蔵し、外部からはM（≦N－2）台のディスクドライブ容量として取り扱われるディスク・アレイ・システムにおいて、

故障のディスクドライブの台数をNkとし、動作可能なディスクドライブの台数をNe（N－Nk）としたときに、「Ne≧M＋2」の条件では、

少なくとも1台以上のディスクドライブを予備ディスクドライブとして停止状態に移行させて、

前記予備ディスクドライブが動作状態に移行することなく、所定時間を経過する毎にアイドリング動作を実行させ、

かつ前記故障のディスクドライブを停止状態に移行させるときに、前記予備ディスクドライブを動作状態に移行させると共に、故障したディスクドライブの全データを前記予備ディスクドライブに復元する処理を実行するコントローラ手段を具備したことを特徴とするディスク・アレイ・システム。

【請求項6】 前記コントローラ手段は、「M≦Ne≦M＋1」の条件では、残る全ディスクドライブを動作状態にする処理を実行することを特徴とする請求項5記載のディスク・アレイ・システム。

【請求項7】 複数のディスクドライブを内蔵して、各ディスクドライブを駆動制御するディスク・アレイ・システムであって、

前記複数のディスクドライブを2群以上のグループに分割し、

ディスクドライブを所定の時間周期で動作状態と停止状態とを交互に繰り返すように制御するコントローラ手段を具備したことを特徴とするディスク・アレイ・システム。

【請求項8】 前記コントローラ手段は、所定の時間周期で動作状態と停止状態とを交互に繰り返す2群のディスクドライブ間において、一方の群が停止状態に移行する前に、停止状態の他方の群のディスクドライブを動作状態に移行して前記一方の群のディスクドライブのデータを前記他方の群のディスクドライブに受け渡すように制御することを特徴とする請求項7記載のディスク・アレイ・システム。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】本発明は、複数のディスクドライブを並列に駆動制御するディスク・アレイ・システムに関する。

【0002】

【従来の技術】従来、大記憶容量と共に、高信頼性に対する要求の強い大規模システムに適用する記憶システムとして、冗長性を有するRAID（redundant arrays of inexpensive disks）装置またはディスク・アレイ・システムと称する磁気ディスクシステムが開発されている。

【0003】このようなRAIDの採用により、システムの信頼性を向上させることが可能となったが、長期的には装置寿命の点で必ずしも十分ではなく、ある期間を経過すると信頼性が急激に低下する恐れがある。

【0004】即ち、RAIDは、複数のハードディスクドライブ（HDD）を内蔵しており、各HDDの保証寿

命が通常では5年程度であるため、5年を経過してから信頼性の問題が発生することになる。実際には、各HDDの寿命にはばらつきがあるため、保証期間を経過を過ぎると、ドライブを構成する部品の磨耗等を要因として故障するHDDが次々と発生することになる。

【0005】従って、高信頼性を得るために複数のHDDを組み合わてRAIDを構成しても、通常では5年以上が経過すると、統計的には半数以上のHDDに寿命が来て、故障したHDDの交換が必要になる。このため、例えば工場施設等で、10年、20年のように連続して長期間の使用を前提としている場合、5年毎に運転を中止してHDDを交換することになる。

【0006】しかしながら、通常では5年も経過すると、RAIDのディスクドライブとして採用したHDDは陳腐化して、生産中止になっている場合も多い。そこで、事前に交換用のHDDを準備しておくことになるが、5年間のような長期間の保存を行なうと、HDDは正常に動作しない場合が多い。これは、例えば磁気ヘッドがディスク上に吸着するような状態になったり、ディスクを回転させるスピンドルモータやヘッドアクチュエータのグリースが固化して、正常に機能しない状態が発生するためである。

【0007】

【発明が解決しようとする課題】前述したように、RAID装置は、複数のHDDを組み合わせて高信頼性の記憶システムを構築できるが、その寿命は各HDDの寿命に依存する5年程度が限度である。このため、長期間の連続使用に耐えるRAIDシステムを維持するには、交換用の予備のHDDを準備するか、交換時に製造元からHDDを供給してもらう必要がある。しかしながら、予備のHDDを準備する方法は、前述したような原因により、使用時に正常に機能しない場合がある。また、製造元からHDDを供給してもらう方法は、5年以上経過していると交換用のHDDが旧型になるため、いわば特注の製品供給となり、コスト負担が大きくなる。

【0008】本発明の目的は、構成する各ディスクドライブの寿命期間を越える長期間の使用を可能にして、最小限のコスト負担で各HDDの寿命に依存しない高信頼性のディスク・アレイ・システムを提供することにある。

【0009】

【課題を解決するための手段】本発明の第1は、複数のディスクドライブを内蔵するディスク・アレイ・システムであって、少なくとも1台以上のディスクドライブを予備ディスクドライブとして停止状態にして、停止状態の開始から時間経過を監視し、所定の時間が経過する毎に停止状態のディスクドライブをアイドリング動作状態に移行させるコントローラ手段を備えたシステムである。コントローラ手段は、所定の条件で動作状態のディスクドライブと予備ディスクドライブとを交換する制御を実行する。このようなシステムにより、寿命がきたディスクドライブと交換可能な予備ディスクドライブを準備すると共に、予備ディスクドライブを所定の間隔でアイドリング動作状態にすることにより、使用時に機能不全となるような事態を防止することができる。

【0010】具体的には、N（≧3）台以上のディスクドライブを内蔵し、外部からはM（≦N－2）台のディスクドライブ容量として取り扱われるディスク・アレイ・システムにおいて、故障のディスクドライブの台数をNkとし、動作可能なディスクドライブの台数をNe（N－Nk）としたときに、「Ne≧M＋2」の条件では、少なくとも1台以上のディスクドライブを予備ディスクドライブとして停止状態に移行させて、前記予備ディスクドライブが動作状態に移行することなく、所定時間を経過する毎にアイドリング動作を実行させ、かつ前記故障のディスクドライブを停止状態に移行させるときに、前記予備ディスクドライブを動作状態に移行させると共に、故障したディスクドライブの全データを前記予備ディスクドライブに復元する処理を実行するシステムである。

【0011】ここで、所定時間はこれ以上ディスクドライブを停止状態に保持しておくと、装置寿命を低下させる要因に基づいて設定される。例えば、ヘッドがディスクに吸着する事を防止するために設定する時間である。即ち、これ以上ヘッドとディスクを接触保持すると吸着の危険が生ずるであろう時間に安全係数を見込んで短めの時間を設定することが好ましい。通常では、その所定時間は、スピンドルモータやアクチュエータに使っているボールベアリングのグリースが固化し、正常に動作しなくなる時間より短時間である。

【0012】本発明の第2は、複数のディスクドライブを2群以上のグループに分割し、少なくとも1群のディスクドライブを所定の時間周期で動作状態と停止状態とを交互に繰り返すように制御するディスク・アレイ・システムである。さらに、一方の群が停止状態に移行する前に、停止状態の他方の群のディスクドライブを動作状態に移行して、一方の群のディスクドライブのデータを他方の群のディスクドライブに受け渡すように制御する。

【0013】具体的には、例えば2群のディスクドライブにより構成されているシステムでは、動作中の一群のディスクドライブがβ時間動作後停止状態になる前に、停止中の他群のディスクドライブにデータを受け渡す時間をαとする。一群のディスクドライブの動作時間は（β＋α）時間となり、他群のディスクドライブの停止時間は（β－α）時間となる。このようにして、2群のディスクドライブはそれぞれ、2β時間周期で動作状態と停止状態とが繰り返されることになる。また、3群のディスクドライブにより構成されているシステムでは、1群のディスクドライブが動作しているときに、データ

の受け渡し時間以外は他の2群のディスクドライブを停止状態にする。このとき、各群のディスクドライブにおいて、動作時間は（β＋α）時間となり、停止時間は（2β－α）時間となる。このようにして、3群のディスクドライブはそれぞれ、3β時間周期で動作状態と停止状態とが繰り返されることになる。

【0014】

【発明の実施の形態】以下図面を参照して本発明の実施の形態を説明する。

（第1の実施形態）図1は第1の実施形態に関係するディスク・アレイ・システムの要部を示すブロック図であり、図2と図3は本実施形態の動作を説明するためのフローチャートであり、図4～図7は本実施形態の動作を説明するための概念図である。

（システム構成）本実施形態のシステム11は、図1に示すように、例えばレベル3のRAIDを想定しており、コントローラ（ディスク・アレイ・制御装置）12、および複数のハードディスクドライブ（以下HDDと称する）1～10を有する。コントローラ12は、各HDD1～10を駆動制御し、ホストコンピュータ13との間でデータの授受を行なうためのインターフェース機能を有する。

【0015】各HDD1～10において、HDD（PD）5は、パリティ情報を保存している専用ドライブである。このPD5に保存されたパリティ情報に基づいて、コントローラ12は故障したHDD（D2X）2の全データを復元する機能を有する。なお、HDD（D2S）6は故障したHDD（D2X）2の代替HDDであることを意味し、またHDD（D7N～D10N）7～10は停止状態（休止状態）のHDDであることを意味する。

【0016】なお、本実施形態は、前記のようにパリティ情報専用のHDD（PD）5を有するレベル3を想定して説明するが、パリティ情報を各HDDに分散させるレベル5の場合でも同様に適用することができる。

（本実施形態の故障発生処理）以下図2のフローチャートを参照して本実施形態の動作について説明する。

【0017】本実施形態では、システムのHDDの全台数をN（ここでN＝10）としたとき、外部からはM（≦N－2）台のHDDを有するRAIDシステムとして取り扱われるものと想定する。コントローラ12は、M＋1番目のHDD5までの各HDD1～5を動作状態にして稼働させ、残りのHDD6～10を休止状態にさせる（ステップS1）。

【0018】コントローラ12は、休止状態の開始から時間経過を監視し、予め設定した所定時間（Ta）が経過したときに、休止状態の各HDD6～10をアイドリング動作状態にさせる（ステップS2のYES、S3）。ここで、アイドリング動作状態とは、実際のデータ記録再生動作時と同様に、スピンドルモータによりディスクを定常回転速度で回転させて、ヘッドアクチュエータを駆動してヘッドをディスクの最内周から最外周まで数回乃至数十回シークさせる動作である。また、所定時間（Ta）は、HDDを停止状態に保持しているときに、装置寿命を低下させる要因の内容に基づいて設定する。例えば、ヘッドがディスクに吸着する事を防止するために設定する時間である。即ち、これ以上ヘッドとディスクを接触保持すると吸着の危険が生ずるであろう時間に、安全係数を見込んで短めの時間を設定することが好ましい。通常では、所定時間（Ta）は、スピンドルモータやヘッドアクチュエータに使っているボールベアリングのグリースが固化し、正常に動作しなくなる時間よりも短時間に設定される。また、ヘッドをロード／アンロードする方式（ランプロード方式）のHDDの場合には、例えば前記グリースの固化時間を目安にして設定することが望ましい。

【0019】ここで、稼働中のHDD1～5の中で、1台のHDD（D2X）2が故障した場合を想定すると、コントローラ12は、休止状態のHDD6～10の中で例えばHDD6を代替HDD（D2S）として起動させる（ステップS4のYES、S5）。このとき、コントローラ12は、HDD（PD）5に保存されているパリティ情報（冗長データ）を利用して、故障したHDD（D2X）2の全データを復元して、代替HDD6に格納する（ステップS6）。

【0020】以下同様にして、つぎの故障HDDが発生したときに、休止状態のHDD7～10から代替HDDを設定して交換させる処理を行なう。この交換処理は、休止状態のHDDから代替HDDが無くなるまで実行される。このような方式であれば、故障したHDDが発生したときに、例えば5年程度の寿命期間に相当する期間が経過している場合、新規のHDDの供給が困難であっても、予め準備した休止状態のHDDから代替HDDと交換することができる。この代替HDDは、前記のように、所定時間Taの間隔でアイドリング動作を繰り返すため、5年程度が経過してから使用する場合でも、直ちに正常な動作状態に移行することが可能である。これにより、寿命により故障したHDDを正常な代替HDDに交換することにより、5年を越える長期間でも高信頼性を維持することができる。

（本実施形態の交互運転方式）次に、図3のフローチャートと図4～図7を参照して、本実施形態の各HDD1～10において、動作状態と休止状態とを交互に行なう方式について説明する。ここでは、図1に示すシステムにおいて、HDD（D2X）2も正常動作のHDDであると想定し、初期時にはHDD6～10を休止状態のHDDとして想定する。従って、初期時には、外部からはHDD1～5によりRAIDシステムが構成されているように見える。

【0021】コントローラ12は時間経過を監視し、所

定時間（Tb）が経過すると、休止状態のHDD6を起動させて、稼働中のHDD1を休止状態にさせる（ステップS12～14）。このとき、コントローラ12は、稼働中のHDD1を休止状態にさせる前に、HDD1に保存されている全データを起動させたHDD6に転送する（ステップS13）。この場合、稼働中のHDD1を休止状態にして、HDD（PD）5に保存されているパリティ情報（冗長データ）を利用して、HDD1の全データを復元して、HDD6に格納してもよい。

【0022】以下、図4から図7を参照して具体的に説明する。所定時間（Tb）をα時間としたときに、図4に示すように、システムの初期状態から、休止状態のHDD6が起動した状態に移行する。ここで、HDD1からHDD6に全データを転送（コピー）に要する時間をβ時間とすると、（α＋β）時間の経過後に、HDD1は休止状態に移行する。以下同様の動作を順に繰り返すと、一連のHDD1～10をサイクリックに稼働させると、定常状態における1周期は10＊（α＋β）時間となる。なお、休止状態のHDDは、前記の所定時間Ta毎にアイドリング動作を実行しても良い。

【0023】ここで、特定の1台の連続動作時間は｛5＊（α＋β）＋β｝時間となり、停止時間は｛5＊（α＋β）－β｝時間となる。通常では、「α>>β」とすれば平均稼働率は約50％となる。即ち、一般化した場合に、定常状態における1周期では、2（M＋1）＊（α＋β）時間のうち特定の1台の連続動作時間は｛（M＋1）＊（α＋β）＋β｝時間となり、停止時間は｛（M＋1）＊（α＋β）－β｝時間となる。通常α>>βとすれば平均稼働率は約50％となる。また、3（M＋1）台により構成されているシステムでは、（M＋1）台のHDDが動作しているとき、他のHDDは停止していることになり、α>>βとすれば特定のドライブの平均稼働率は約33％となる。

【0024】また、バッファメモリを使用することにより、故障したHDDの全データの復元中を含めたHDDの交換作業中においても、システムを連続使用することができる。休止状態にさせるべきHDDが無くなった場合には、当然ながら残る全HDDを動作状態にさせる。さらに、HDDの連続運転の上限を決める前記所定時間（Tb）が前記所定時間（Ta）より短時間に設定すれば、アイドリング時間を設ける必要が無くなるので好適である。HDD間のデータ転送処理に実用上の支障にならなければ、「Tb≦Ta」と設定してもよい。

【0025】本実施形態は、HDDを1台ずつを順番に交換させる場合ついて説明したが、一度に交換させるHDDは複数であっても良い。1台のHDDを順に交換させる場合に比べて、一般にバッファメモリを初めとするハードウェア量が増えるとともにデータ書換の時間が長くなるが、前記の所定時間Tb，Taにおいて「Tb＝Ta」の条件を満たすことができる場合は、アイドリングが不要になるという特長を有する。

【0026】次に、本実施形態の応用例として、図5に示すように、故障したHDD2が発生した場合に、代替HDD6を使用した修復処理後に、故障したHDD2に続くHDD3から、前記の休止状態とデータの復元処理を実行する方式もある。即ち、修復処理後のα時間後から一連のHDD1～10をサイクリックに稼働させると、定常状態における1周期は17＊（α＋β）時間となる。また、図6に示すように、故障したHDD2が発生した場合でも、本来のHDD1からの順番で動作状態と休止状態の各HDDを交換する方法でもよい。この場合、HDD1の次は故障したHDD2の代替HDD6となる（2（α＋β）時間後）。一方、このような順番とは全く関係無く、連続稼働時間の長いものから順に交換させる方法でもよい。

【0027】さらに、図7に示すように、HDD1～5とHDD8～10のディスクドライブ群によりRAIDシステムを構成し、残りの2台のHDD6，7をミラーディスク装置（M1，M2）として構成する方式を提案する。この方式であれば、アクセス頻度の高いデータをミラーディスク装置（M1，M2）に格納し、それ以外のデータをHDD1～5に振り分けることができる。従って、HDD1～5の記録時のオーバーヘッドを軽減化することができる。また、高速の画像データをRAIDに格納し、それ以外のデータをミラーディスク装置（M1，M2）に格納してもよい。本実施形態では、ミラーディスク装置（M1，M2）として設定したHDD6，7を除く残りのHDD1～5とHDD8～10に対して、休止状態と動作状態を順番に交換させる。要するに、この方式であれば、本実施形態の利点とミラーディスク装置（M1，M2）としての利点の両方を得ることが可能である。

【0028】以上のような本実施形態の動作を一般化した場合に、K台のHDDをサイクリックに稼働させると、定常状態における1周期は、K（α＋β）時間となり、このうち特定の1台の連続動作時間は｛（M＋1）＊（α＋β）＋β｝時間となり、停止時間は｛（K－M＋1）＊（α＋β）－β｝時間となる。従って、HDDの平均稼働率（％）は｛（M＋1）（α＋β）＋β｝／｛K（α＋β）｝＊100（％）となる。即ち、特定のHDDの平均稼働率（D）は「D＝（M＋1）（α＋β）＋β｝／｛K（α＋β）｝」となる。単独HDDの平均寿命をLとすると、各HDDの平均寿命Ltは1／D倍となり、「Lt＝L／D＝L＊（α＋β）／（M＋1）（α＋β）＋β｝」となる。本実施形態のように10台のHDDからなるシステムであれば、外部からは4台分のHDD容量を有するようなシステムとした場合に、「α＋β」を24時間，βを0．5時間とした場合を例にとれば、各HDDの平均寿命は9．96年となり、ほぼ2倍の寿命となる。

（第2の実施形態）次に、図8と図9を参照して第2の実施形態について説明する。第2の実施形態は、2群以上のHDD21A，21Bに分割したシステム（図8）、または2群以上のRAID31～33からなるシステム（図9）を想定し、各HDDまたは各RAIDを交互に動作状態と停止状態を繰り返す方式である。

【0029】具体的には、図8に示すシステムにおいて、コントローラ20は、例えば2台のHDD21A，21Bそれぞれ、交互に動作と停止を繰り返すように制御する。このとき、コントローラ20は、HDD21A，21Bを例えば24時間の並列動作を実行し、次にHDD21Bのみ24時間の停止状態とする。次に、HDD21Bを動作させて、HDD21AからHDD21Bにデータを転送させる。

【0030】この間に、ホストコンピュータからコントローラ20に対して、データ出力（データの読出し）の指令があったときには、コントローラ20はHDD21Aからデータを出力する。また、データ入力（データの書き込み）の指令があったときには、コントローラ20はHDD21Bにデータを入力する。データの授受が終了した時点で、コントローラ20はHDD21Aを停止状態に移行させて、HDD21Bのみを動作させる。

【0031】そして、HDD21Bの動作状態で24時間が経過すると、コントローラ20はHDD21Aを起動して、HDD21BからHDD21Aにデータを転送させる。以下同様の動作を繰り返して、HDD21AとHDD21Bを交互に動作させることにより、RAIDシステム全体の寿命は、個々のHDDの寿命のおよそ2倍となる。

【0032】また、HDD21AとHDD21B間のデータの転送時間は短いほど、システムの寿命は長くなるため、全データを受け渡すのではなく、書き換えられたデータのみ受け渡すことの方が良い。HDD21AとHDD21Bのいずれもが動作中で故障すると、最新データは失われるが、停止状態のHDDがバックアップとして機能するため、故障する1ステップ前のデータは確保することができる。従って、前記のように初期時に、HDD21AとHDD21Bを並列に動作させるのはバックアップを取るためである。なお、HDD21AとHDD21B間で、データ受け渡し中に、ホストコンピュータから入力されたデータは最新データとして保護するため、データの受け渡し処理を禁止する。

【0033】また、本実施形態の応用例として、図9に示すように、3群のRAID31～33からなるシステムであり、各RAID31～33が並列に同時動作するミラーディスク構成のHDD31A，31B，HDD32A，32B，およびHDD33A，33Bを有するシステムを想定している。

【0034】このようなシステムにおいて、コントローラ30は、1群のRAID31が例えば24時間動作しているときに、他の2群のRAID32，33が48時間停止するように、それぞれ交互に動作と停止とを繰り返す制御を行なう。これにより、システムの各RAIDの寿命に対して、約3倍に寿命を延ばすことが可能となる。各RAID31～33をそれぞれ、ミラーディスク構成にすることにより、故障によりデータを消失する確率を大幅に低下させることができる。また、各RAID31～33は動作と停止が間欠的に起きるため、磁気ヘッドは定期的にディスク上のCSS（Contact Start Stop）エリアに移動することになる（HDDの停止時）。従って、ヘッドはCSSエリアに接触するため、ごみ等の汚れが取れて、連続浮上しているときに生じやすいヘッドクラッシュを防止することもできる。

【0035】なお、交互に停止と動作を繰り返すディスクシステムは、単体の装置でも良いし、RAIDシステムでもよい。RAIDシステムでは、故障時にデータが破壊されることはない。

【0036】以上のような本実施形態の内容を一般化すると、1群のHDDまたはRAIDの動作時間を（$\beta+\alpha$）時間とし、他群の停止時間を（$\beta-\alpha$）時間とすると、動作と停止をこのように繰り返すことにより、一周期2β時間中に個々の装置（HDDまたはRAID）は（$\beta+\alpha$）時間しか動作しないことになる。このため個々の装置の寿命をL時間とすると、システム全体の寿命Ltは、「$Lt=L*2\beta/(\beta+\alpha)$」となる。ここで、$\beta$＝24時間、$\alpha$＝0．5時間、L＝5年、のとき、システムとしての寿命Ltは9．8年となる。

【0037】また、3群のHDDまたはRAIDにより構成されているシステムでは、動作時間は（$\beta+\alpha$）時間、停止時間は（$2\beta-\alpha$）時間とすると、個々の装置寿命をLとするシステム全体の装置寿命Ltは、「$Lt=L*3\beta/(\beta+\alpha)$」となり、L＝5年、$\beta$＝24時間、$\alpha$＝0．5時間では、Lt＝14．7年となり、結果的に長寿命を得ることができる。

【0038】

【発明の効果】以上詳述したように本発明によれば、第1に常に予備のディスクドライブを準備し、予備のディスクドライブを定期的に停止状態と動作状態とを交互に繰り返すことにより、長期間の経過後でも正常に動作させることを実現して、結果的にシステム全体の寿命を個々のディスクドライブの寿命より延ばすことができる。従って、ディスクドライブの交換が必要となときに、特注の製品供給等をなくすことが可能であるため、最小限のコスト負担で長期間の連続使用に耐える高信頼性のRAIDシステムを提供することができる。また、冗長性を持たせる2群以上のディスクドライブ（RAIDも含む）を動作状態と停止状態とを繰り返すことにより、前記と同様に、システム全体の寿命を個々のディスクドライブの寿命より延ばすことができる。従って、結果的

に、最小限のコスト負担で、各ＨＤＤの寿命に依存しない高信頼性のＲＡＩＤシステムを構築することが可能である。

【図面の簡単な説明】

【図１】本発明の第１の実施形態に関係するディスク・アレイ・システムの要部を示すブロック図。

【図２】第１の実施形態の動作を説明するためのフローチャート。

【図３】第１の実施形態の動作を説明するためのフローチャート。

【図４】第１の実施形態の動作を説明するための概念図。

【図５】第１の実施形態の動作を説明するための概念図。

【図６】第１の実施形態の動作を説明するための概念図。

【図７】第１の実施形態の動作を説明するための概念図。

【図８】第２の実施形態に関係するシステムの構成を示すブロック図。

【図９】第２の実施形態に関係するシステムの構成を示すブロック図。

【符号の説明】

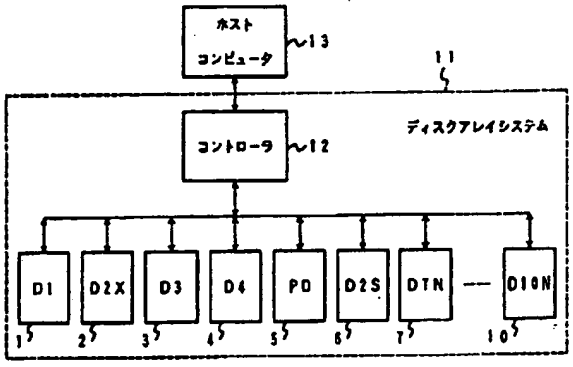
１～１０，２１Ａ，２１Ｂ…ディスクドライブ（ＨＤＤ）

１１，３１，３２，３３…ＲＡＩＤシステム

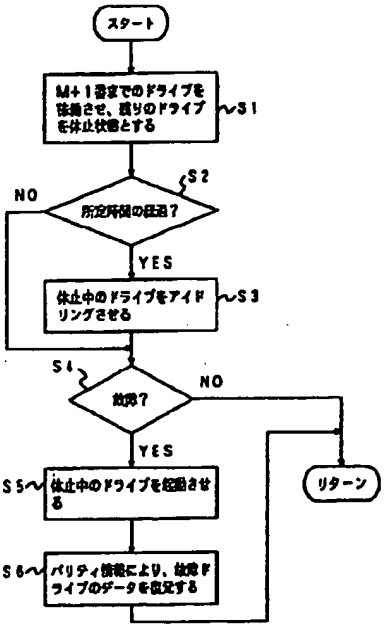
１２，２０，３０…コントローラ

１３…ホストコンピュータ

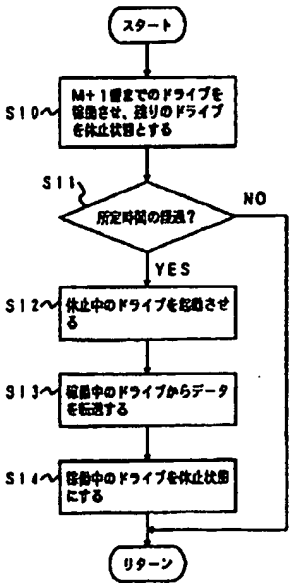
【図１】



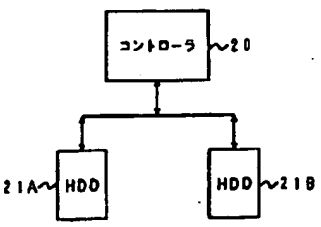
【図２】



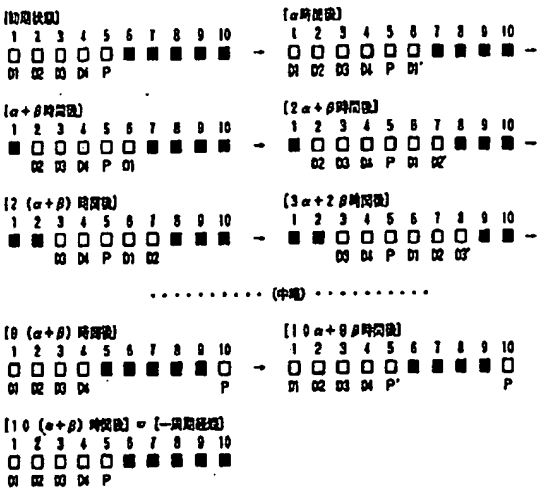
【図３】



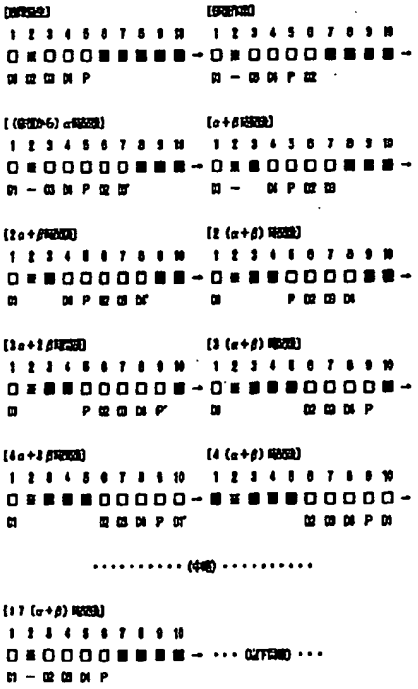
【図８】



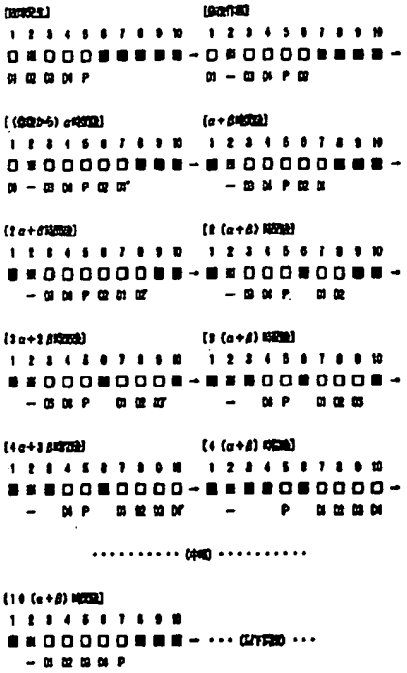
【図4】



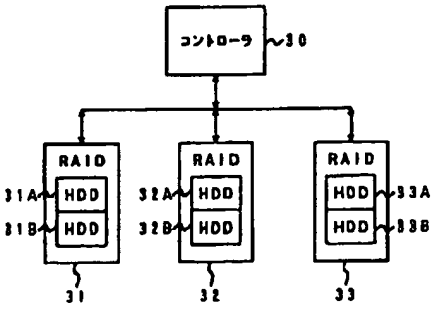
【図5】



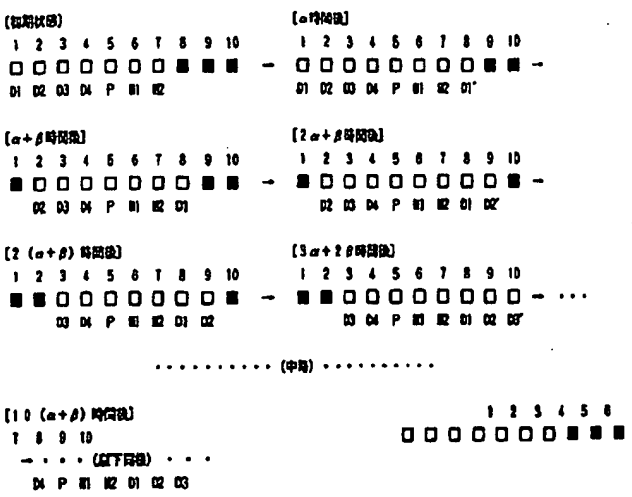
【図6】



【図9】



【図7】


(初期状態)
1 2 3 4 5 6 7 8 9 10
D1 D2 D3 D4 P B1 B2
[αT時後]
1 2 3 4 5 6 7 8 9 10
D1 D2 D3 D4 P B1 B2 D1'
[α+βT時後]
1 2 3 4 5 6 7 8 9 10
D2 D3 D4 P B1 B2 D1
[2α+βT時後]
1 2 3 4 5 6 7 8 9 10
D2 D3 D4 P B1 B2 D1 D2'
[2(α+β)T時後]
1 2 3 4 5 6 7 8 9 10
D3 D4 P B1 B2 D1 D2
[3α+2βT時後]
1 2 3 4 5 6 7 8 9 10
D3 D4 P B1 B2 D1 D2 D3'
(中略)
[10(α+β)T時後]
7 8 9 10
(以下同様)
D4 P B1 B2 D1 D2 D3
1 2 3 4 5 6